Northeastern University, Boston, MA

College of Engineering

Department of Civil and Environmental Engineering

CIVE 7381: Transportation Demand Forecasting and Model Estimation

Problem Set 3

From: **Nathan David Obeng-Amoako** (NUID: 002607282)

To: Professor Haris Koutsopoulos

Submitted on: Friday, October 25th

Fall 2024

Table of Contents

List of Figures

List of Tables

List of Equations

List of DataFrames

List of Abbreviations

OLS = Ordinary Least Squares

RMSEE = root mean square error

TAZ = traffic analysis zone

Problems

Northeastern

Department of Civil and
Environmental Engineering
400 Snell Engineering
360 Huntington Ave.
Boston, MA 02115

CIVE 7381 Transportation Demand
Problem Set #4
Due: Monday, October 28, 2024

Problem 1

The file *Boston_TAZ_Data.xlsx* (Canvas, under PS 4) contains data from more than 2,720 TAZs (traffic analysis zones) in the Boston Metropolitan area. For each zone you have data on the total number of HBW trips **attracted** in the zone, in addition to land use information related to the zone. You will use the data to develop, using linear regression, a model that predicts the trips/day attracted by the zone.

a) Briefly discuss the cause/effect relationships you think are relevant for modeling trip attractions.

b) Present relevant statistical summaries and descriptive statistics on the data provided. Check for the presence of outliers. Also detect any other issues with the data.

c) Experiment with different linear regression models that can predict the number of trips attracted in each zone. You can use any software you are familiar with (excel, R, python, SAS, etc.) Report the results of your best model. Explain your choice by evaluating and validating each model, based on various statistical tests discussed in class (coefficient of determination, t-statistics, F-statistic confidence interval). Also discuss and interpret the model specification and the impact of the various parameters. Use scatter plots and other visual tools to illustrate your model performance (e.g., scatter plots of actual vs model predicted values).

d) Do the data and model results support the relationships you propose in a)?

The data file contains the following variables that can be used for the specification of your model. Use the variables as is or transform them to capture the effects that you think are important.

ID: ID of each TAZ
HBW_a: number of home-based work trip attractions in the TAZ
Total_Emp: total employment
Srv_Emp: service employment
Ret_Emp: retail employment
Bas_Emp: basic employment
K12_Emp: K12 private and public school employment
Coll_Emp: College employment
HH: number of households in the TAZ

Problem 2

An initial (seed) trip matrix is available. The trip matrix is symmetric and shown below. The table also shows the total (target) number of trips produced or attracted in each zone for the future planning year.

From\To	Origin OD matrix 1	2	3	4	5	6	Target Row Total (Productions)
1	5	40	120	30	50	60	800
2		5	52	55	60	100	400
3			10	25	90	30	400
4				10	15	45	200
5					15	55	500
6						20	700
Target Column Total (Attractions)	200	600	400	300	1000	500	3000

a. Using the above seed trip matrix estimate the 6x6 trip matrix for the target year. Iterate until it converges to an F value (target/predicted for each zone productions and attractions) between 0.90 to 1.10 (but no more than 3 iterations). Keep track of each row's and column's factor. Report your results for two different approaches and compare the two outputs:

   i. Average factor
   ii. Biproportional fitting IPF)

b. The travel time between the various zones is given below. Find the matrix of basic friction values for each zonal pair, using an exponential friction function, $f(t_{ij}) = exp(\beta t_{ij})$, with parameter β = -0.05.

From\To	Travel Times (min) 1	2	3	4	5	6
1	10	25	37	40	40	50
2	25	10	52	35	30	40
3	37	52	10	55	40	50
4	40	35	55	10	25	37
5	40	30	40	25	10	25
6	50	40	50	37	25	10

c. With the final trip distribution (OD matrix) found in question a) part i) above, calculate the average travel time (HINT: it should be a weighted average).

d. Apply the gravity model **once** (no balancing) for different values of the parameter β (assuming the productions, attractions, travel times and friction function given earlier). Assume that β values vary from -0.02 to -0.12 in increments of 0.02. Plot the average trip time against the corresponding value of the parameter β. Comment on the relationship.

e. Assume that the planning authority has estimated that the true average trip time is 26 minutes. Based on the results from part d) above, choose the value of the parameter β that is the most appropriate.

Problem 3

A small area is divided into 6 zones. The travel times t_{ij} between the zones are given in the table below (the travel times are symmetric):

Zone	1	2	3	4	5	6
1		10	17	16	12	22
2			7	6	15	12
3				13	8	5
4					21	9
5						13
6						

The target productions and attractions for each zone are given below:

Zone	P's	A's
1	1000	100
2	1500	500
3	100	3000
4	200	500
5	1700	50
6	780	1130
Total	5280	5280

a. Find the friction factors for each OD pair assuming that the friction function is given by:

$$f(t_{ij}) = \frac{1}{t_{ij}^2}$$

b. Find the zone to zone trip flows using the gravity model. Use the iterative updating of attractions and repeated application of the gravity model (discussed in class) to find the target trip interchanges (do not use the bi-proportional method to balance the attractions and productions). Perform no more than 2 iterations. You may use the practical considerations discussed in class to approximate the **intrazonal** travel times.

Solution to Problem 1

Likely Predictor Variables

First of all, land use plays a critical role in trip attraction. An industrial area is expected to attract less trips than a commercial area with a bunch of retail establishments because the latter attracts a bunch of consumer trips. The employment dynamics (types and density) of the area may also impact trip attraction. For instance, areas with high concentration of service and retail employment can attract more trips. Moreover, the transportation accessibility of the area can impact trip attraction. For example, if there's no public transportation in the area and there are inadequate bicycle facilities, then, most likely, only people who are privileged to own a car will be attracted to the area. Lastly, temporal factors like time of day or even season of year can impact trip attraction rates. Other factors like household size and density may also affect trip attraction rates.

Statistical Summaries and Descriptive Statistics

Descriptive statistics of the relevant features in the dataset are shown in ***DataFrame 1.***

```
dF4.drop(columns=['Zone_ID']).describe()
```

	HBW_a	Total_Emp	Srv_Emp	Ret_Emp	Bas_Emp	K12_Emp	Coll_Emp	HH
count	2727.000000	2727.000000	2727.000000	2727.000000	2727.000000	2727.000000	2727.000000	2727.000000
mean	1113.883343	878.082508	558.323799	155.811881	190.861019	45.297396	37.695636	657.966997
std	1388.731627	1124.752877	958.279870	242.731325	392.887037	81.933835	257.651080	423.375389
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	343.974626	275.500000	141.000000	21.000000	25.000000	0.000000	0.000000	338.500000
50%	685.879005	547.000000	295.000000	70.000000	63.000000	0.000000	0.000000	663.000000
75%	1346.539581	1030.500000	593.000000	192.000000	174.500000	67.000000	0.000000	949.500000
max	15074.108240	12953.000000	12975.000000	3016.000000	7652.000000	647.000000	7225.000000	2547.000000

DataFrame 1: Descriptive Statistics of Relevant Features

From ***DataFrame 2***, there are no missing data in the dataset.

```
# Missing values
dF4.isnull().sum().to_frame().transpose()
```

	Zone_ID	HBW_a	Total_Emp	Srv_Emp	Ret_Emp	Bas_Emp	K12_Emp	Coll_Emp	HH
0	0	0	0	0	0	0	0	0	0

DataFrame 2: Check for Missing Values

The boxplots in **Figure 1** show the distribution of the data in the relevant columns. From the figure, it can be seen that all the employment variables, viz. **Total_Emp**, **Srv_Emp**, **Ret_Emp**, **Bas_Emp**, **K12_Emp**, and **Coll_Emp** are very highly positively skewed, which means that most of the TAZs have a relatively low number of employments while a few of them (like the Central Business Districts) have a very high concentration of jobs. It is also observed that **HBW_a** is also highly positively skewed (perhaps, as a result). The number of households per TAZ (**HH**) is approximately symmetric even though it has a few outliers at the right tail. I choose not to delete the outliers as they may represent genuine, valuable information rather than errors or noise. I don't want to lose meaningful data. Also, I tested and found out that deleting them, based on the chosen criterion, will rather screw things up (e.g., with half of the data being lost) or will have very insignificant impact on the results.

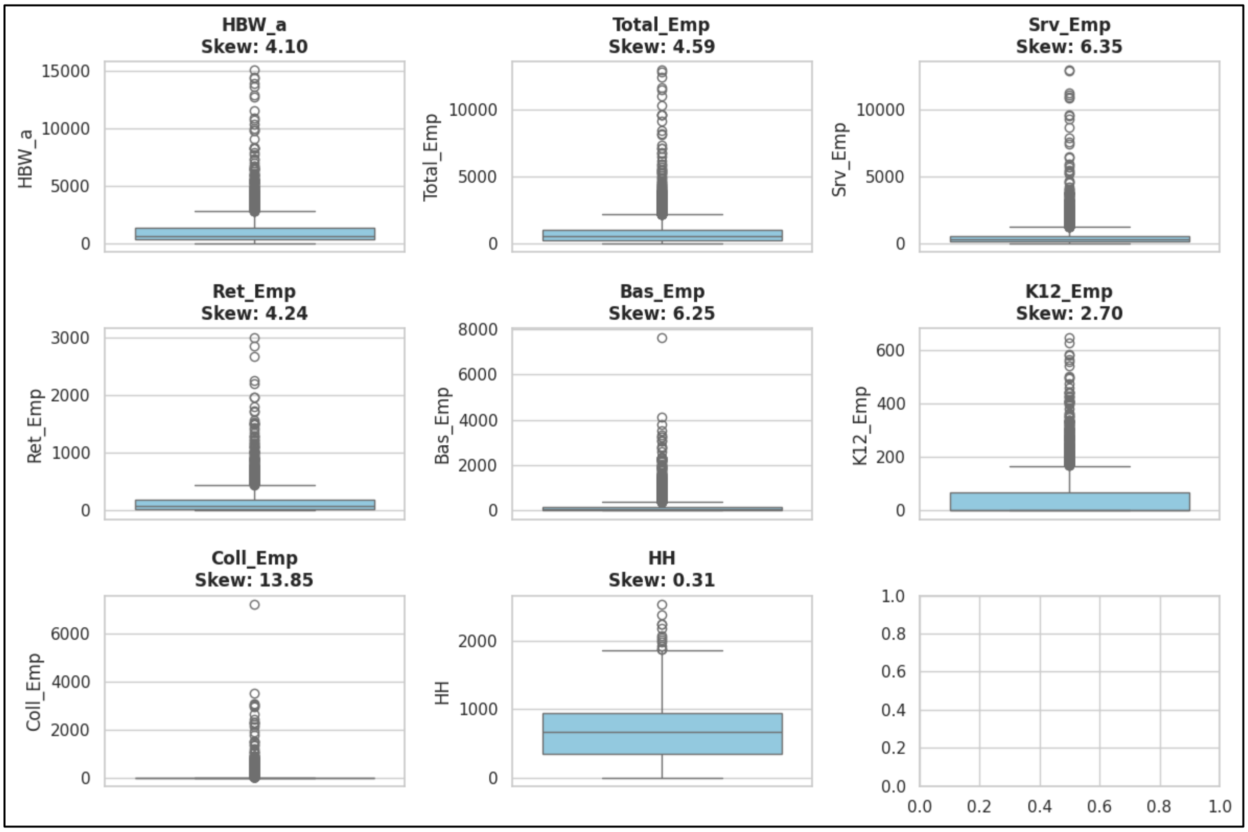

Figure 1: Boxplots of the Relevant Features

Figure *2* shows the histogram of the relevant columns, and this provided me with additional insights on the distribution of the data.

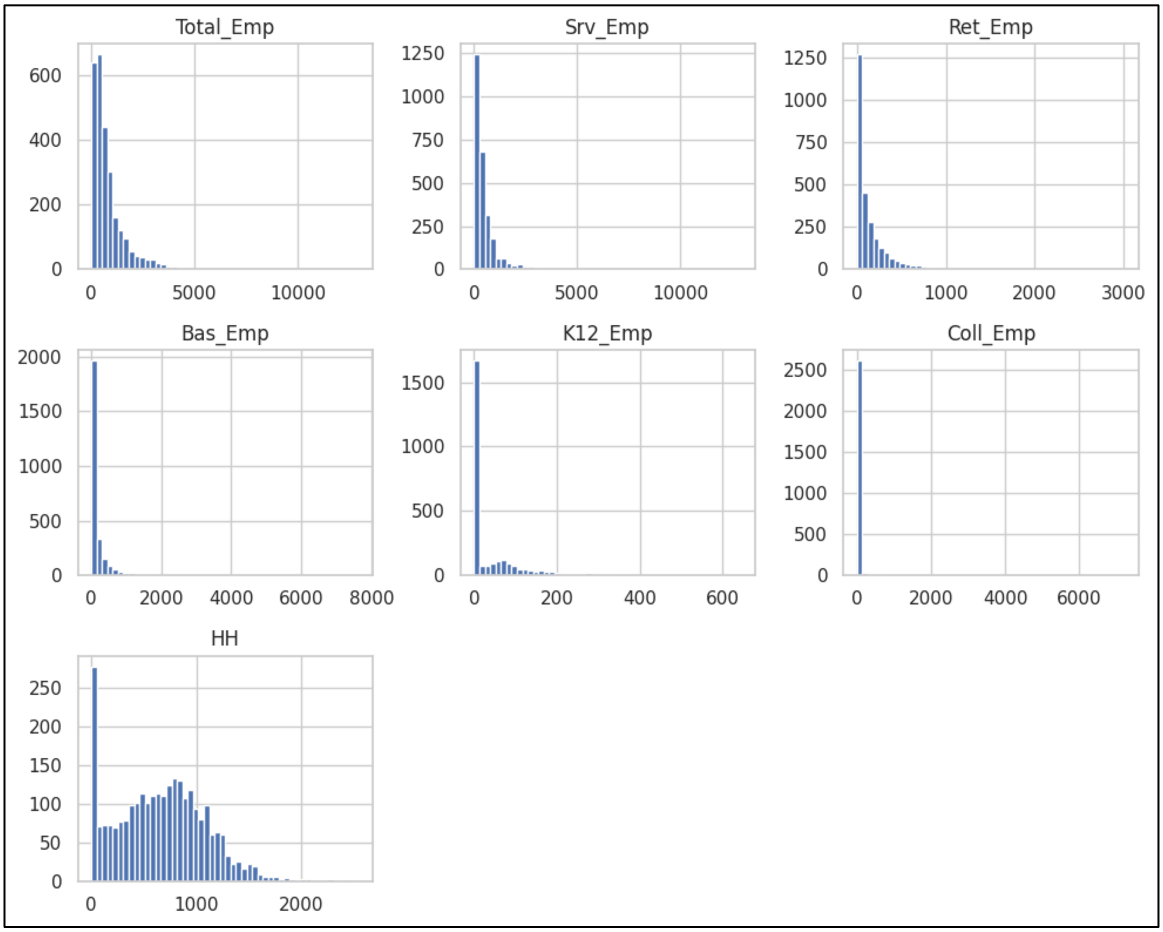

Figure 2: Histograms of the Relevant Features

Models Development

The heatmap to visualize the correlation matrix is illustrated in **Figure 3**.

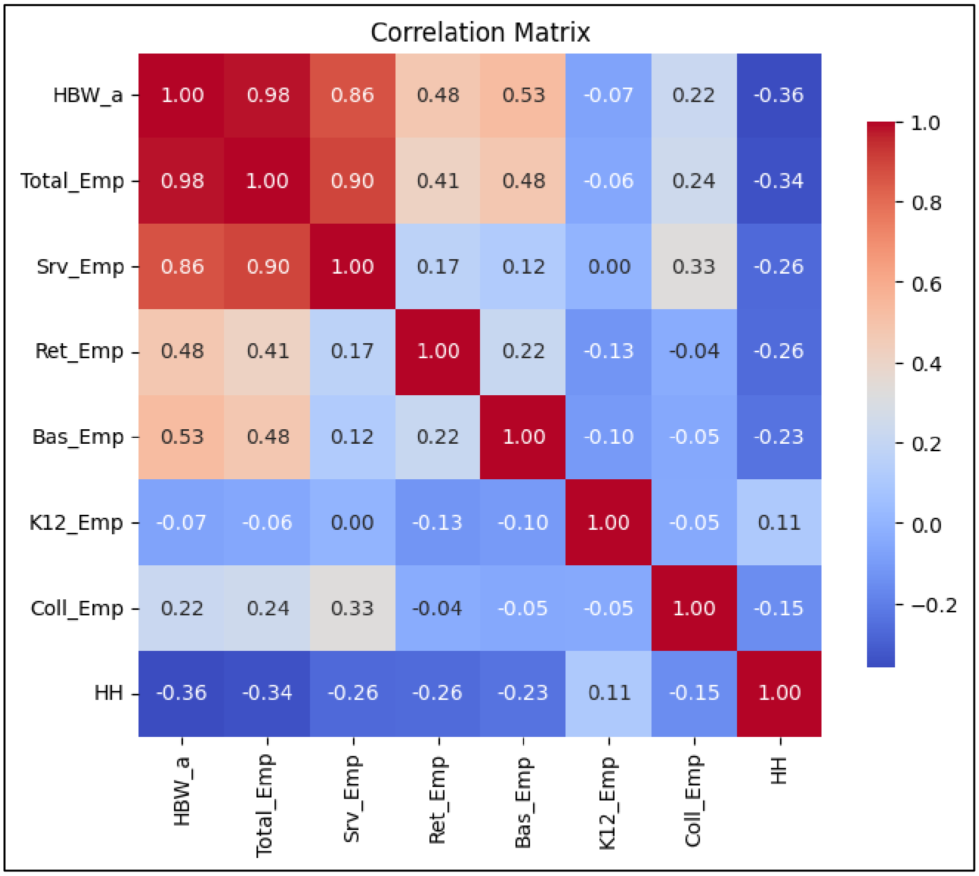

Figure 3: Correlation Heatmap

I tested a variety of models but will focus on the top three best performers. The models with interaction terms did not improve performance, so I decided to exclude them. For instance, I examined several interaction terms derived from **Total_Emp** and **Srv_Emp**, which show strong multicollinearity (R = 0.90). However, these models did not yield better results, prompting me to prioritize a more parsimonious approach.

Models Tested

Model 1 was created using **all** the predictor variables. Model 2 included only those predictor variables with a correlation of **at least 0.5** with the response variable (rounded to one place of decimal). Model 3 focused **solely on the variable Total_Emp**, which shows a very strong correlation with the response variable (0.98) according to the correlation matrix.

Model 1: $$HBW\_a = \beta_0 + \beta_1 \times Total\_Emp + \beta_2 \times Srv\_Emp + \beta_3 \times Ret\_Emp + \beta_4 \times Bas\_Emp + \beta_5 \times K12\_Emp + \beta_6 \times Coll\_Emp + \beta_7 \times HH + \varepsilon$$

Model 2: $$HBW\_a = \beta_0 + \beta_1 \times Total\_Emp + \beta_2 \times Srv\_Emp + \beta_3 \times Ret\_Emp + \beta_4 \times Bas\_Emp + \varepsilon$$

Model 3: $$HBW\_a = \beta_0 + \beta_1 \times Total\_Emp + \varepsilon$$

Coefficients of Determination (R^2)

Table *1* shows a summary of the Coefficients of Determination results for all the three models. The full results are shown in the **Appendix**.

Table 1: Coefficients of Determination Results

	Model 1	**Model 2**	**Model 3**
R^2	0.995	0.995	0.965
Corrected R^2	0.995	0.995	0.965

Based on the R^2 results, I prefer Model 2 over Model 1, as both models explain roughly 99.5% of the variance in the dependent variable, **HBW_a**. However, Model 2 is more parsimonious and Model 1 may be overfitting. Model 3 is performing relatively poorly.

Predicted vs. Actual

Figure *4* illustrates the relationship between predicted and actual values of the response variable for each of the top three models.

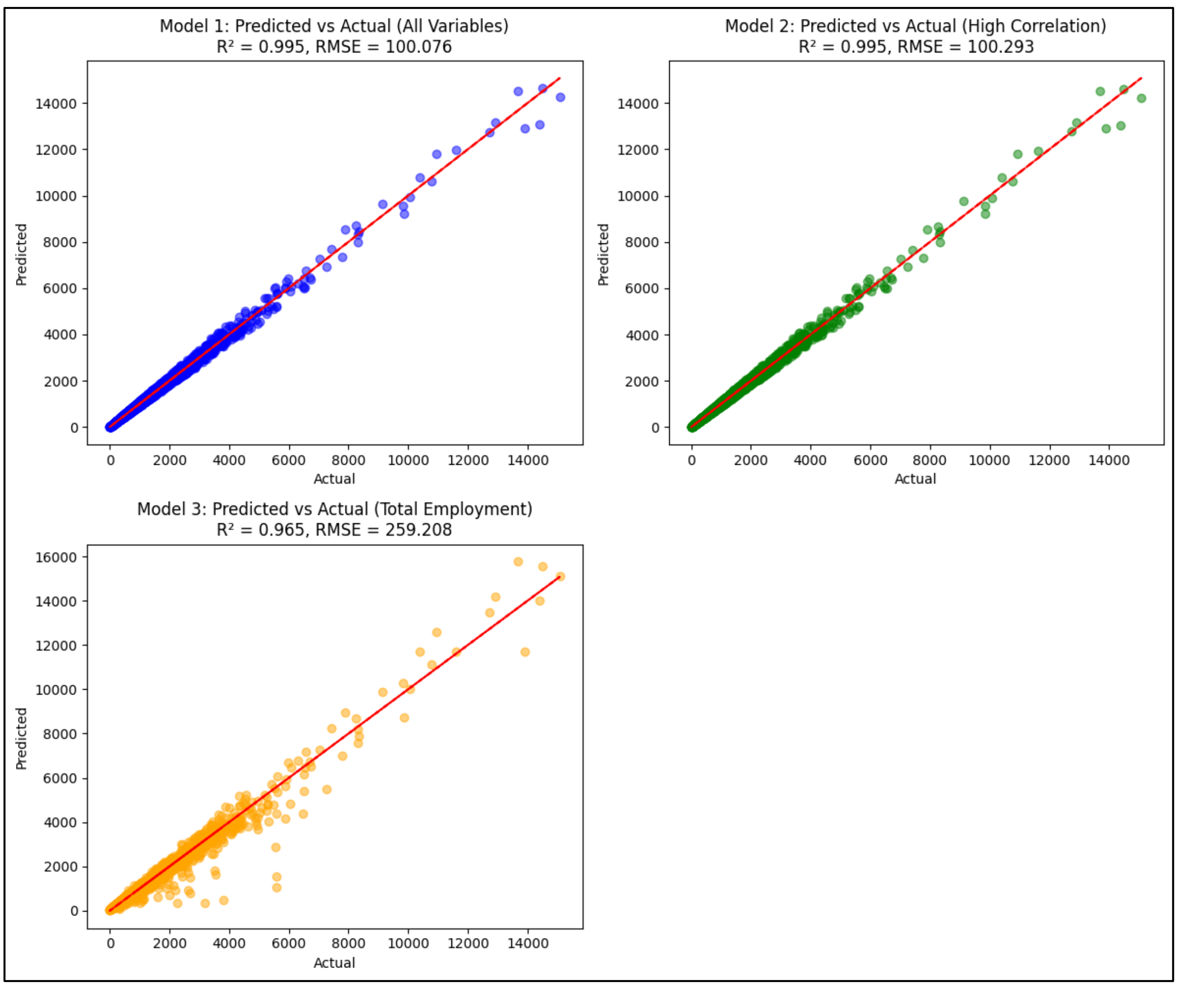

Figure 4: Predicted vs. Actual

As expected, Model 3 exhibits the highest RMSE (from **Figure 4**). Model 1 has a slightly lower RMSE than Model 2, but this difference is negligible and does not provide sufficient reason to prefer Model 1 over Model 2. Additionally, Model 1 may be at risk of overfitting.

T-statistics

Based on the t-statistics values (**Appendix**), it can be concluded that the additional variables in Model 1 are not contributing that much to the model. They're only making it complex in a sense.

For the various reasons stated above, I nominate **Model 2** as the chosen model!

Verification of Initial Hypotheses

For the most part, YES, the data and model results support the proposed relationships.

Solution to Problem 2

Trip Distribution Methods

Average Factor Method

The average factor method adjusts a trip matrix by calculating row and column factors based on the ratio of target productions and attractions to total current trips. It iteratively updates the matrix by applying the average of these factors until the predicted trips converge closely to the targets.

Results after first iteration

	1	2	3	4	5	6	Predicted Pi	Target P	Fi
1	8.2	90.9	230.8	64.3	153.3	127.1	305.0	800.0	2.622951
2	38.8	8	65.1	81.1	143.7	144.7	312.0	400.0	1.282051
3	112.7	81.8	12.2	36.1	212.9	42.5	327.0	400.0	1.223242
4	26.5	83.4	29.2	13.9	34.6	61.3	180.0	200.0	1.111111
5	60.3	110.3	134	25.7	39.5	92.6	285.0	500.0	1.754386
6	87.4	209.1	52.2	88.3	158.6	38.7	310.0	700.0	2.258065
Predicted Aj	305.0	312.0	327.0	180.0	285.0	310.0	nan	nan	nan
Target A	200.0	600.0	400.0	300.0	1000.0	500.0	nan	nan	nan
Fj	0.655738	1.923077	1.223242	1.666667	3.508772	1.612903	nan	nan	nan

Results after second iteration

	1	2	3	4	5	6	Predicted Pi	Target P	Fi
1	7.3	100.7	225	69.3	194.1	138	674.6	800.0	1.185885
2	27.7	7.4	51.9	73	156.5	131.5	481.5	400.0	0.830792
3	79	74.9	9.6	32	228.8	38.1	498.4	400.0	0.802597
4	18.6	76.4	22.9	12.3	37.2	54.8	249.0	200.0	0.803372
5	50.6	116.4	123.6	26.3	47.9	95.7	462.3	500.0	1.081544
6	74.4	222.8	48.8	91.5	194.3	40.4	634.3	700.0	1.103588
Predicted Aj	333.9	583.6	523.5	309.4	742.7	507.0	nan	nan	nan
Target A	200.0	600.0	400.0	300.0	1000.0	500.0	nan	nan	nan
Fj	0.599051	1.028174	0.764039	0.969584	1.346495	0.986259	nan	nan	nan

Results after third iteration

	1	2	3	4	5	6	Predicted Pi	Target P	Fi
1	6.8	105.3	215.9	71.9	218.7	144.4	734.4	800.0	1.089298
2	23.1	7.1	44.7	68.5	160.9	124.6	448.1	400.0	0.892728
3	64.8	69.9	8.1	29.6	232.2	35.5	462.3	400.0	0.865148
4	15.6	72.7	19.8	11.6	38.4	52.2	222.3	200.0	0.899824
5	47.1	121.5	118.4	27.2	53.9	100	460.6	500.0	1.085470
6	67.6	227.7	45.6	92.7	214.3	41.3	672.3	700.0	1.041243
Predicted Aj	257.7	598.6	481.8	304.5	858.8	498.6	nan	nan	nan
Target A	200.0	600.0	400.0	300.0	1000.0	500.0	nan	nan	nan
Fj	0.776203	1.002338	0.830227	0.985176	1.164386	1.002801	nan	nan	nan

Final Trip Matrix (Rounded After Iteration 3)

	1	2	3	4	5	6
1	7	105	216	72	219	144
2	23	7	45	69	161	125
3	65	70	8	30	232	36
4	16	73	20	12	38	52
5	47	121	118	27	54	100
6	68	228	46	93	214	41

Notes:

The convergence criteria were not quite met even after the third iteration. However, I experimented and found out that the criteria will be met after the fifth iteration.

Biproportional Fitting Method

The biproportional method, also known as the Iterative Proportional Fitting (IPF) method, adjusts a trip matrix to match specified row and column totals through iterative proportional adjustments. This technique iteratively modifies the entries of the matrix until the row and column sums converge to their respective target values, maintaining proportional relationships between the data.

Results after first iteration

	1	2	3	4	5	6	Predicted Pi	Target P	Fi
1	5.6	111	194.6	74	270.2	153.1	808.6	800.0	0.989420
2	21.9	6.8	41.2	66.3	158.5	124.7	419.5	400.0	0.953527
3	62.8	67.3	7.6	28.8	226.8	35.7	428.9	400.0	0.932533
4	14.3	64.7	17.2	10.5	34.3	48.6	189.5	200.0	1.055301
5	37.5	111.4	97.6	24.8	54.2	93.9	419.3	500.0	1.192347
6	57.9	238.9	41.9	95.6	255.9	43.9	734.2	700.0	0.953483
Predicted Aj	200.0	600.0	400.0	300.0	1000.0	500.0	nan	nan	nan
Target A	200.0	600.0	400.0	300.0	1000.0	500.0	nan	nan	nan
Fj	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	nan	nan	nan

Results after second iteration

	1	2	3	4	5	6	Predicted Pi	Target P	Fi
1	5.5	108.4	186.3	74.5	274.2	148.8	797.9	800.0	1.002693
2	20.9	6.4	38	64.3	155	116.9	401.5	400.0	0.996206
3	58.5	61.9	6.8	27.3	217	32.7	404.3	400.0	0.989421
4	15	67.3	17.5	11.2	37.2	50.4	198.8	200.0	1.006252
5	44.7	131.1	112.6	30	66.3	110	494.7	500.0	1.010619
6	55.2	224.8	38.6	92.7	250.3	41.2	702.8	700.0	0.995953
Predicted Aj	200.0	600.0	400.0	300.0	1000.0	500.0	nan	nan	nan
Target A	200.0	600.0	400.0	300.0	1000.0	500.0	nan	nan	nan
Fj	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	nan	nan	nan

<u>Final Trip Matrix (Rounded After Iteration 2)</u>

	1	2	3	4	5	6
1	6	108	186	74	274	149
2	21	6	38	64	155	117
3	59	62	7	27	217	33
4	15	67	18	11	37	50
5	45	131	113	30	66	110
6	55	225	39	93	250	41

Notes:

Amazingly, the convergence criteria were met after the second iteration so I didn't even bother to do the third iteration. It's interesting to me how the IPF method is much more efficient than the average factor method. Since they both take just about the same efforts to develop, I guess I'm not ever using the average factor method in real-life application.

Friction Factor Matrix

$f(t_{ij}) = \exp(\beta t_{ij})$, with $\beta = -0.05$

	Zone 1	Zone 2	Zone 3	Zone 4	Zone 5	Zone 6
Zone 1	0.606531	0.286505	0.157237	0.135335	0.135335	0.082085
Zone 2	0.286505	0.606531	0.074274	0.173774	0.223130	0.135335
Zone 3	0.157237	0.074274	0.606531	0.063928	0.135335	0.082085
Zone 4	0.135335	0.173774	0.063928	0.606531	0.286505	0.157237
Zone 5	0.135335	0.223130	0.135335	0.286505	0.606531	0.286505
Zone 6	0.082085	0.135335	0.082085	0.157237	0.286505	0.606531

Average Travel Time

The average (weighted) travel time calculated using **Equation 1** is 36.26 minutes.

$$\mathrm{AWTT} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{m} (F_{ij} \times T_{ij})}{\sum_{i=1}^{n} \sum_{j=1}^{m} T_{ij}}$$ Equation 1: Average Travel Time

Where:

$AWTT$ = average weighted travel time

F_{ij} = friction factor from Zone i to Zone j

T_{ij} = number of trips from Zone i to Zone j

m, n = dimensions of the matrices

Gravity Model

The gravity model applied for the trip distribution; is shown in **Equation 2**.

$$T_{ij} = \frac{P_i A_j F_{ij}}{\sum_j P_i A_j F_{ij}}$$ Equation 2: Gravity Model

Where:

P_i = total productions at Zone i

A_j = total attractions at Zone j

F_{ij} = friction factor from Zone i to Zone j

T_{ij} = number of trips from Zone i to Zone j

The gravity model results for each specified value of β is shown in **Figure 5**.

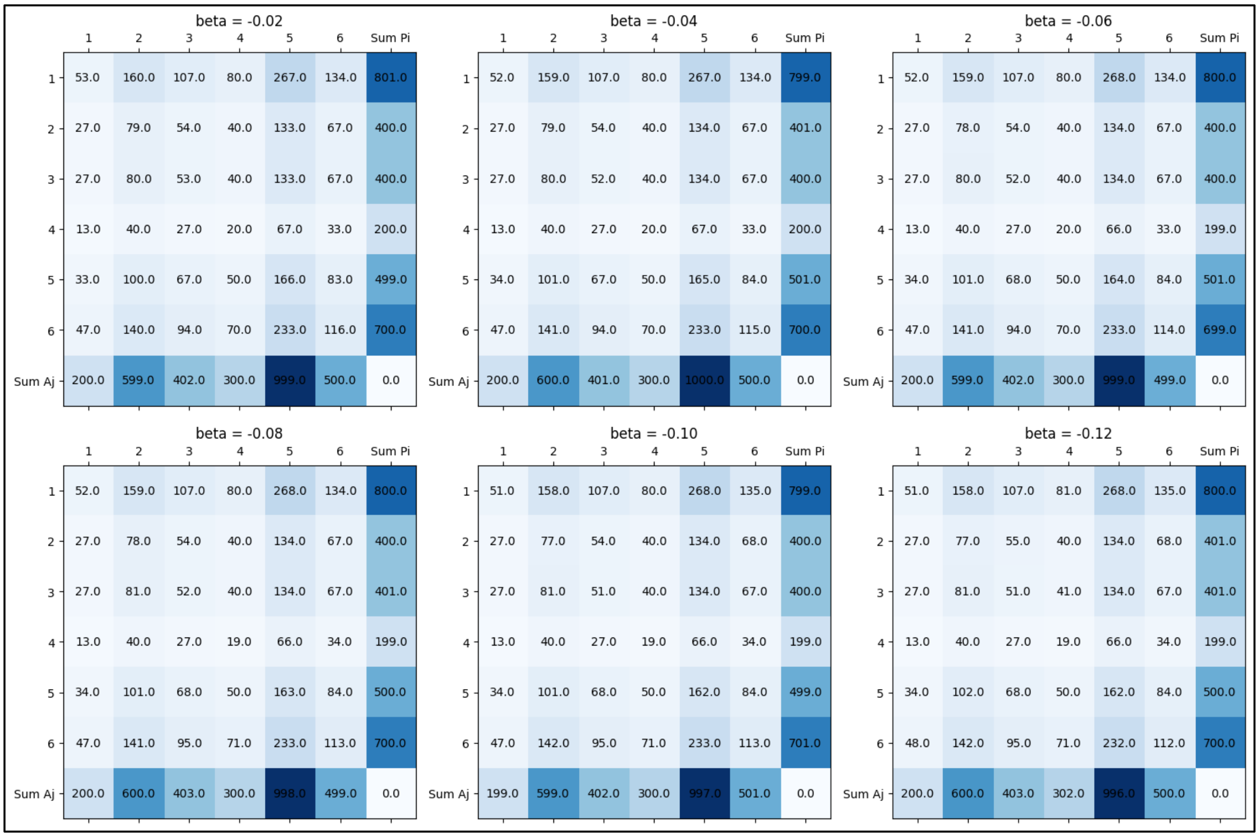

Figure 5: Gravity Model Results

Average Trip Time Against Beta

The plot of average trip time against the corresponding value of β is shown in **Figure *6***. Amongst the chosen β values, the travel time is closest to 26 minutes for β = -0.02. Looking at the trend, it can be assumed that even bigger values of β can result in a travel time that's even much closer to 26 minutes.

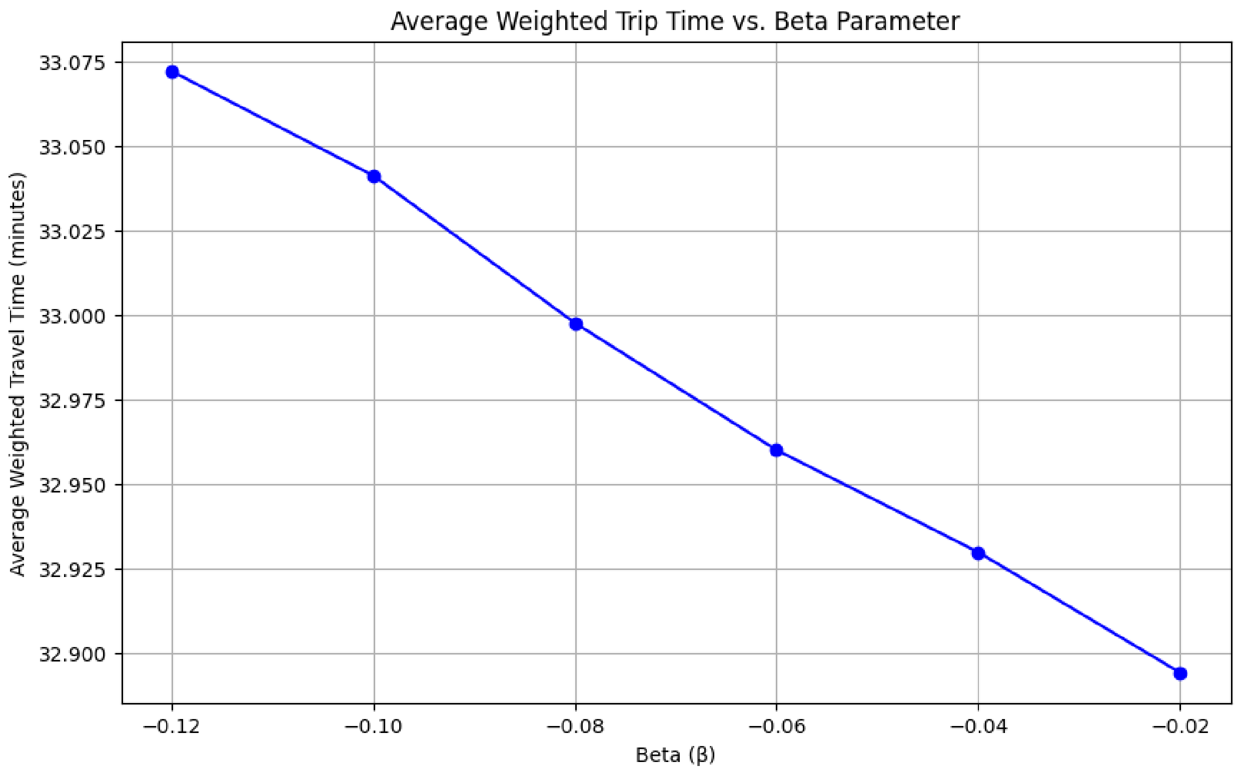

Figure 6: Average Trip Time against Beta

Problem 3

Friction Factors

The friction factor matrix is displayed in **DataFrame *3***. The intrazonal travel time was taken to be half of the average travel time to the nearest zone, which explains why the leading diagonal doesn't have zeros.

	1	2	3	4	5	6
1	0.040000	0.010000	0.003460	0.003906	0.006944	0.002066
2	0.010000	0.111111	0.020408	0.027778	0.004444	0.006944
3	0.003460	0.020408	0.160000	0.005917	0.015625	0.040000
4	0.003906	0.027778	0.005917	0.111111	0.002268	0.012346
5	0.006944	0.004444	0.015625	0.002268	0.062500	0.005917
6	0.002066	0.006944	0.040000	0.012346	0.005917	0.160000

DataFrame 3: Friction Factor Matrix

Gravity Model

The first iteration for the zone-to-zone trip flows using the gravity model (**Equation 2**) is shown in **DataFrame *4***.

	1	2	3	4	5	6	Total Pi
1	166.558000	208.197000	432.244000	81.327000	14.458000	97.216000	1000.000000
2	10.734000	596.353000	657.205000	149.088000	2.385000	84.235000	1500.000000
3	0.064000	1.891000	88.973000	0.548000	0.145000	8.378000	100.000000
4	0.769000	27.327000	34.926000	109.307000	0.223000	27.448000	200.000000
5	19.437000	62.199000	1312.013000	31.734000	87.468000	187.149000	1700.000000
6	0.518000	8.710000	301.015000	15.484000	0.742000	453.530000	780.000000
Total Aj	198.080000	904.677000	2826.376000	387.489000	105.421000	857.957000	0.000000

DataFrame 4: Zone-to-Zone Trips after Iteration 1

For the second iteration, the attractions were updated using the formula in the screenshot illustrated in **Figure 7** (taken from the Professor's notes).

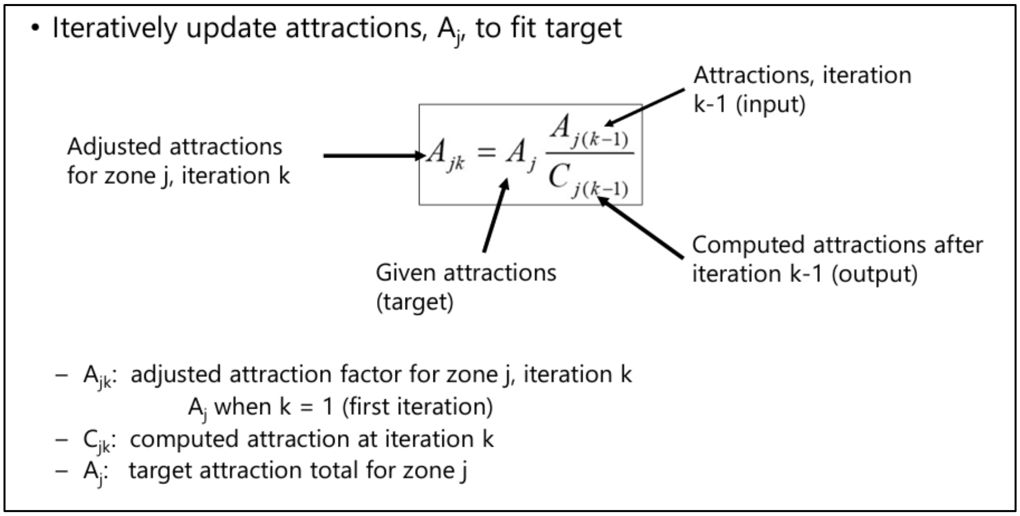

Figure 7: Formula to Update Attractions for Gravity Model

The second iteration for the zone-to-zone trip flows is shown in **DataFrame 5**.

	1	2	3	4	5	6	Total Pi
1	93.659000	128.167000	511.029000	116.888000	7.638000	142.619000	1000.000000
2	6.080000	369.765000	782.597000	215.824000	1.269000	124.466000	1500.000000
3	0.030000	0.974000	87.991000	0.659000	0.064000	10.282000	100.000000
4	0.338000	13.141000	32.255000	122.720000	0.092000	31.454000	200.000000
5	9.448000	33.097000	1340.774000	39.424000	39.941000	237.316000	1700.000000
6	0.217000	3.985000	264.489000	16.540000	0.291000	494.478000	780.000000
Total Aj	109.771000	549.128000	3019.135000	512.056000	49.295000	1040.615000	0.000000

DataFrame 5: Zone-to-Zone Trips after Iteration 2

Appendix

MODEL 1 – USING ALL VARIABLES

```
Model 1: Using all variables
                            OLS Regression Results
==============================================================================
Dep. Variable:                  HBW_a   R-squared:                       0.995
Model:                            OLS   Adj. R-squared:                  0.995
Method:                 Least Squares   F-statistic:                 7.438e+04
Date:                Fri, 25 Oct 2024   Prob (F-statistic):               0.00
Time:                        18:37:15   Log-Likelihood:                -16430.
No. Observations:                2727   AIC:                         3.288e+04
Df Residuals:                    2719   BIC:                         3.292e+04
Df Model:                           7
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         -9.6297      4.784     -2.013      0.044     -19.011      -0.248
Total_Emp      0.0091      0.012      0.788      0.431      -0.014       0.032
Srv_Emp        1.1134      0.011     97.505      0.000       1.091       1.136
Ret_Emp        1.5163      0.014    111.185      0.000       1.490       1.543
Bas_Emp        1.3266      0.012    108.562      0.000       1.303       1.351
K12_Emp        0.0168      0.024      0.707      0.480      -0.030       0.064
Coll_Emp      -0.0226      0.008     -2.815      0.005      -0.038      -0.007
HH             0.0068      0.005      1.379      0.168      -0.003       0.017
==============================================================================
Omnibus:                      973.274   Durbin-Watson:                   2.067
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            70391.086
Skew:                           0.808   Prob(JB):                         0.00
Kurtosis:                      27.837   Cond. No.                     4.54e+03
==============================================================================
```

MODEL 2 – USING HIGH CORRELATION VARIABLES

```
Model 2: Using high correlation variables
                            OLS Regression Results
==============================================================================
Dep. Variable:                  HBW_a   R-squared:                       0.995
Model:                            OLS   Adj. R-squared:                  0.995
Method:                 Least Squares   F-statistic:                 1.297e+05
Date:                Fri, 25 Oct 2024   Prob (F-statistic):               0.00
Time:                        18:37:15   Log-Likelihood:                -16436.
No. Observations:                2727   AIC:                         3.288e+04
Df Residuals:                    2722   BIC:                         3.291e+04
Df Model:                           4
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         -3.4909      2.542     -1.373      0.170      -8.475       1.493
Total_Emp      0.0108      0.012      0.928      0.353      -0.012       0.033
Srv_Emp        1.1091      0.011     97.590      0.000       1.087       1.131
Ret_Emp        1.5138      0.013    112.426      0.000       1.487       1.540
Bas_Emp        1.3246      0.012    108.940      0.000       1.301       1.348
==============================================================================
Omnibus:                     1001.025   Durbin-Watson:                   2.060
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            72902.994
Skew:                           0.857   Prob(JB):                         0.00
Kurtosis:                      28.272   Cond. No.                     2.37e+03
==============================================================================
```

MODEL 3 – USING ONLY TOTAL EMPLOYMENT

```
Model 3: Using only Total Employment
                            OLS Regression Results
==============================================================================
Dep. Variable:                  HBW_a   R-squared:                       0.965
Model:                            OLS   Adj. R-squared:                  0.965
Method:                 Least Squares   F-statistic:                 7.546e+04
Date:                Fri, 25 Oct 2024   Prob (F-statistic):               0.00
Time:                        18:37:15   Log-Likelihood:                -19025.
No. Observations:                2727   AIC:                         3.805e+04
Df Residuals:                    2725   BIC:                         3.807e+04
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         48.7753      6.300      7.742      0.000      36.422      61.129
Total_Emp      1.2130      0.004    274.708      0.000       1.204       1.222
==============================================================================
Omnibus:                     3394.282   Durbin-Watson:                   1.886
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           887193.468
Skew:                           6.426   Prob(JB):                         0.00
Kurtosis:                      90.424   Cond. No.                     1.81e+03
==============================================================================
```